AI-Enhanced Air Crash Investigation and Safety Improvement Classification

Sambhav Jain	Hardik Garg	Kunal Surtani	Ayush Bhakat	Abhinav Nambiar
Reg.No: (220962063)	*Reg.No: (220962071)*	*Reg.No: (220962041)*	*Reg No: (220962089)*	*Reg. No: (220962242)*
Roll.No: 19	Roll.No: 21	Roll.No: 14	Roll.No: 23	Roll.No: 38

***Abstract*—** This paper presents a novel method for enhancing air crash investigations using a combination of AI and Big Data technologies. The system focuses on analyzing the National Transportation Safety Board (NTSB) aviation dataset to classify flights as either safe or unsafe, based on calculated risk percentages. By leveraging the Hadoop framework for distributed data storage, the system ensures efficient handling of large datasets, while Hive manages real-time data streaming and structured querying. The preprocessing phase utilizes PySpark to clean, filter, and transform the data, preparing it for analysis in a scalable and efficient manner. A deep learning model is then applied to the processed data, identifying potential risk factors based on historical flight information, environmental conditions, and various flight parameters. Flights that exceed a specific risk threshold are flagged as unsafe, providing critical insights into possible safety hazards.

***Keywords*—**Hadoop, Hive, PySpark, Air Crash Investigation, Risk Factor Analysis, Machine Learning, Data Processing

I. INTRODUCTION

Air crashes are catastrophic events that result in significant loss of life and resources, making thorough investigations crucial for understanding the causes and preventing future occurrences. Traditionally, the investigation process involves analyzing vast amounts of data, including flight logs, weather conditions, and maintenance records, often done manually. This approach is time-consuming and prone to human oversight. However, advancements in artificial intelligence (AI) and Big Data technologies now offer a more efficient and accurate way to conduct air crash investigations. This project aims to leverage these cutting-edge tools to classify flights based on risk factors, enabling investigators to focus on critical incidents more effectively. By analyzing data from the National Transportation Safety Board (NTSB) aviation dataset, the system provides a detailed risk assessment for each flight, identifying potential safety concerns before they lead to incidents.

At the core of the project is a deep learning model that processes a wide range of parameters to assess risk factors from flight data. This model evaluates variables such as flight history, environmental conditions, and operational patterns to classify flights as either safe or unsafe, assigning a risk score to each. This automated analysis provides a faster and more objective method for identifying patterns and anomalies that could contribute to an accident. By pinpointing high-risk flights, the system helps prioritize investigations, streamlining the decision-making process and enabling a more proactive approach to safety management in aviation.

Literature Review

Reference [1]:

This paper applies a set of data-mining and sequential deep-learning techniques to accident investigation reports published by the National Transportation Safety Board (NTSB) in support of the prognosis of adverse events. It detects patterns in either the event sequences or in the raw text narratives that can be used to predict the possible outcome of a potentially adverse situation. NTSB data extracts are laid out in a relational database format. In this study, the data extract is converted into a MySQL data format using database utilities. This makes the data easy to query and analyze. The Long Short-term Memory (LSTM) was used for training the prognosis model using the word-embedding format of either the raw text input data or the event sequences as input data since LSTM neural network has proven to be efficient at making predictions on ordered sequences of data like time-series data. The model accuracy varies from 72.92% to 81.19% during cross-validation. The model predicts the probability of class label, i.e. whether the event is an accident or incident according to probabilities computed in the output layer of the neural network model.

Reference [2]:
The research predicts the likely causes of airplane crashes using many machine learning algorithms such as XGBoost, CHAID, Tree-AS and deep neural network and compares the results and accuracies among these algorithms. The dataset consists of airplane crashes since 1908 which consists of various features such as Flight Number, Aircraft Type, Fatalities detail and summary of the accident. The causes were classified into seven categories: Weather Error, Air Traffic Control and Navigation Error , Pilot and Staff Error , Unknown , Miscellaneous and Machine Fault. Due to the large size of the dataset, it is divided into 3 parts of approximately 40 years of data and processed individually using the HDFS system in Hadoop and clustered appropriately. XGBoost algorithm proved to be the best machine learning model with accuracy of 59.68% in the training set and 39.52% in the testing set. The average accuracy provided by all the models remained close to 40% which means there is significant room for improvement through added features and limitation of features in the dataset.

Reference [3]:
This paper aims to increase en-route flight safety through the development of deep learning models for trajectory prediction, where model prediction uncertainty is characterized following a Bayesian approach. In the first step, a large volume of raw messages in Flight Information Exchange Model (FIXM) format streamed from the Federal Aviation Administration are processed with a distributed computing engine Apache Spark to extract trajectory information in an efficient manner. As a unified engine for big data analytics, Apache Spark combines a core engine for distributed computing with an advanced programming model for in-memory data processing through a data sharing schema — resilient distributed dataset (RDD). The Spark framework's in-memory programming model results in up to 100 times faster than Hadoop in big data analytics. Two types of Bayesian neural networks are trained to make flight trajectory prediction from different perspectives. In both models, model prediction uncertainty is quantified with the dropout strategy, and propagated through the integrated model with Monte Carlo samples. The two types of Bayesian neural networks are integrated for making accurate long-term predictions for ongoing flights. By doing this, we leverage the advantages of both models: the high prediction accuracy of the DNN model and longer-term prediction capability of the LSTM model. A probabilistic safety indicator is used to measure the horizontal and vertical separation distance between two flights. The LSTM and DNN integrated model makes a much more accurate prediction than the single LSTM model. The integrated model demonstrates a much better performance w.r.t. the common quantities of the two models: latitude and longitude prediction. The value of RMSE is reduced by 49% and 58% compared to the LSTM model, respectively.

Reference [4]:
The NTSB aviation accident database contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters up to February 2021. It is the dataset used in our research project due to its factual correctness and also due to availability of large number of details related to a particular aviation accident in form of 32 features, some of them including Event ID, Event Type, Location, Latitude, Longitude, Aircraft Damage, Make, Model and Total Fatalities. After data cleaning and applying encoding techniques to categorical features, we have performed feature extraction to reduce the dimensionality of the dataset and successfully compute the risk factor of given airplane accidents.

Research Gaps and Objectives

Research Gaps:

1. **Limited Real-Time Analysis**: Most studies focus on retrospective analysis of crash data, which can miss out on real-time insights that could be critical in preventing future incidents. The ability to process live streams of flight data, environmental inputs, and other dynamic factors in real time is underexplored. Real-time AI systems would enable aviation authorities to detect risk factors as they emerge during flights. The current lack of emphasis on this gap limits the proactive capacity of AI models in mitigating accidents before they occur.
2. **Generalization Across Diverse Conditions**: AI models are often trained on specific aviation datasets, which are context-dependent and may not fully capture the diversity of global aviation conditions. For example, variations in weather, flight routes, aircraft models, and maintenance standards across regions are not always considered. As a result, models developed in one context may not generalize well to others, reducing their utility in diverse operational environments. Research should focus on building

more versatile models that can handle a wider range of flight scenarios and conditions to improve robustness and accuracy.

3. **Human Factors and Complex Interactions**: Air crash investigations require a deep understanding of human factors, such as pilot decision-making, crew coordination, and communication under stress. Current AI models excel in analyzing quantitative data but struggle to incorporate the complexities of human behavior and interactions with systems. These factors are often critical in crash causality but are challenging to model accurately. There is a research gap in developing AI models that can better account for these qualitative aspects of flight operations, especially in terms of predicting how human errors may interact with technical failures.
4. **Data Quality and Standardization**: A key issue is the quality and consistency of the data used for training AI models. Incomplete, inconsistent, or non-standardized data from flight logs, maintenance records, and other sources can significantly impair the performance of AI systems. There is currently no global standard for how aviation data is collected and stored, leading to compatibility issues when building comprehensive AI systems that need to process data from various airlines and regulatory bodies. Addressing this gap requires the development of global data-sharing protocols and improving the quality of data inputs for AI models.

Objectives:

1. **Develop a Comprehensive Classification Model for Investigation Type Using Hadoop Ecosystem**: Utilize PySpark to build a scalable machine learning pipeline that can preprocess large datasets and encode categorical vehicle features. Apply clustering and deep learning techniques to classify air crash investigation into types such as Accident or Incident.
2. **Identify and Analyze Feature Interactions in Air Crash Investigation**: Explore how various airplane features and flight details contribute to the overall risk and acceptability score by analyzing their interactions. Investigate methods to quantify these relationships to reveal how factors like aircraft damage, injury severity, the number of uninjured passengers, and weather conditions collectively impact classification decisions in air crash investigations. This analysis can uncover the combined effects of multiple variables, providing deeper insights into the factors that determine the type and depth of investigations required for different incidents.
3. **Incorporate Deep Learning for Enhanced Classification Accuracy**: Explore the use of deep learning techniques, such as artificial neural networks, transformers for improving classification accuracy. By leveraging TensorFlow and neural network architectures, the model can better capture complex patterns in vehicle data that traditional methods may overlook.
4. **Enhance Model Interpretability**: Focus on improving the interpretability of the classification models by incorporating techniques like feature importance analysis, backtracking and model visualization. Provide stakeholders with clear insights into how different features contribute to classification decisions, enhancing the model's usability for real-world applications.
5. **Evaluate Model Performance Using Robust Metrics**: Evaluate the classification models based on accuracy, precision, recall, and F1-score to ensure reliable performance. Benchmark these metrics against baseline models and validate the efficacy of both clustering and deep learning approaches in accurately classifying air crash incidents.
6. **Address Real-World Deployment Considerations for Air Crash Investigation:** To deploy AI models in real-world air crash investigations, models must be computationally efficient and scalable, particularly given the vast amount of aviation data involved. Optimizing for speed is crucial to allow rapid risk assessment of incoming flight data in near real-time. Models should also be robust enough to handle diverse data sources such as flight logs, maintenance records, and environmental data streams. Additionally, maintaining flexibility for future updates is essential as aviation technologies and safety standards evolve, ensuring the model's relevance and effectiveness over time.

Methodology

Data Collection and Preprocessing

4.1.1 Data Acquisition

- The dataset was obtained from Kaggle , consisting of **97417 rows** and **97 columns**. The dataset is checked for null values and the total number of null values present in each column. Subsequently, some columns are removed from the dataset which have a number of null values greater than a particular threshold (70000). Additional columns are dropped through manual feature elimination based on intuition that the feature may not be important to predict the type of air crash involved. The features remaining in the dataset include:

- **Make.Model.Name**
- **Mission**
- **Flight.Phase**
- **Problem**
- **Human.Factors**
- **Anomaly**
- **Narrative**

This dataset is essential for understanding the factors influencing vehicle acceptability. We utilized PySpark's capabilities to import the dataset into a DataFrame, allowing for efficient processing and manipulation of large-scale data.

4.1.2 Data Preprocessing

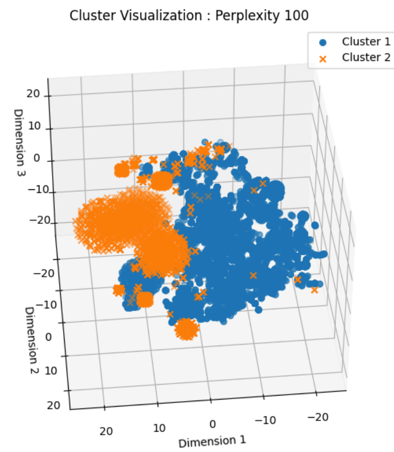
- The dataset underwent a thorough preprocessing phase to ensure compatibility with the classification model:
  - **Encoding Categorical Variables:** Categorical variables were transformed into numerical format using ordinal encoding. This transformation is critical for the classification model to interpret the features accurately. For example, the feature 'Mission' had 4 unique values: ['Personal', 'Passenger', 'Training', 'Test Flight' ,'Cargo'] along with null values in some entries. Thus, the values in the lists were mapped from 0 to 3 respectively in the dataset. Similar encoding was performed on other categorical features.

  - **Handling Missing Values:** The entries with missing values (nan) were ultimately removed from the dataset.

  - **Text Based columns :** The 'Narrative' feature contained text based information of the flight journey. Using nlp techniques such as embeddings , the feature was mapped to a higher dimensional space

  - **KMeans Clustering to obtain classes :** The Dataset consisted of enlisted flight related issues without a clear distinction between the problems stated. Leveraging K means clustering identifies two classes which correlates with a flight problem being 'critical' or 'not critical'

Representing the two classes, the clusters segregate the data efficiently.



KMeans Clustering plot

4.1.3 Feature Scaling

- Feature scaling was applied to standardize the feature values, which mitigates the influence of scale differences among features. We employed the StandardScaler method from PySpark's MLlib. The scaling formula is represented as follows:

$$X_{\text{scaled}} = \frac{X - \text{mean}(X)}{\text{standard deviation}(X)}$$

Where:

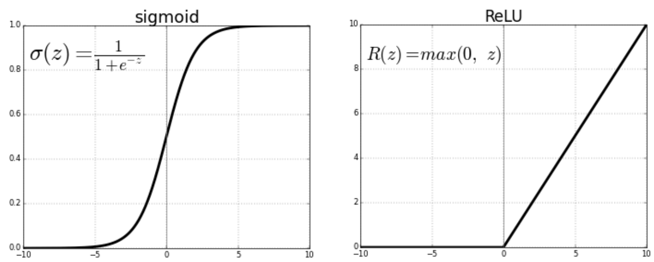
- X represents the original feature values.
- mean(X) is the mean of the feature values.
- standard deviation(X) is the standard deviation of the feature values.

4.2 Model Selection and Training

4.2.1 Model Architecture

- Using the Tensorflow library, we have constructed a modified transformer based on the pretrained 'Bert' model. The approach involves converting the 'narrative' feature into tokens that is parsed through the transformer for feature extraction.

- Simultaneously a separate set of inputs that consists of categorical features that are fed into a deep neural network for feature extraction. Both feature outputs are then combined together and trained using a fully connected Neural Network that performs binary classification. The activation functions used are ReLU (Rectified Linear Unit) and Sigmoid in the Output Layer.



Sigmoid and ReLU functions

4.2.1 Multi Head Attention

- For the predictive model, we selected a combined approach due to its effectiveness in nlp as well as numerical data classification tasks. The model predicts the probability of an air crash disaster being classified into one of the acceptability categories (critical or not critical). The fundamental equation used is the Multi-Head Attention Mechanism. This function can mathematically be represented as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V$$

Where:

- Q is the query matrix,
- K is the key matrix,
- V is the value matrix,
- W_i^Q , W_i^K , W_i^V , and W^O are the learned projection matrices,
- d_k is the dimension of the key vector.

4.2.2 Model Training

- The dataset was divided into training and testing sets using an 80-20 ratio. The combined model was trained on the training set using the fit() method. During training, the model adjusted the coefficients to minimize the log loss, defined as:

$$\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}\left[Y_i \log(\hat{Y}_i) + (1 - Y_i)\log(1 - \hat{Y}_i)\right]$$

Where:

- N is the total number of observations.
- Y_i is the actual class label for observation i.
- $\hat{Y}_i$ is the predicted probability for observation i.

4.3 Model Evaluation

4.3.1 Evaluation Metrics

- The model's performance was evaluated using key metrics such as accuracy, precision, recall, and F1-score. These metrics are essential for assessing the model's predictive capability:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

4.4 Model Deployment and Usage

4.4.1 Model Usage

- The model serves as a tool to predict the type of crash the aircraft is involved in by feeding values of features such as severity of injury, number of uninjured passengers, damage to aircraft and number of engines in aircraft.

- To test the performance of the model on unseen data, we are designing a website which accepts the values of input features which is then linked to the PySpark script for processing and fed to the deep layer neural network to predict the value. Continuing this process for different input samples, we get a gauge of the performance of the model and could change certain parameters and model architecture by comparing the classification metrics before and after the modification.

4.5 Model Maintenance and Updates

4.5.1 Performance Monitoring

- To ensure sustained effectiveness, the deployed model undergoes regular performance monitoring, tracking key metrics such as prediction accuracy and processing latency, enabling prompt detection of any performance deviations.

4.5.2 Model Updates

As new data becomes available or conditions in aviation change, it's crucial to periodically update the AI model to maintain its accuracy and relevance in air crash investigations. This involves retraining the model with fresh datasets, such as updated flight logs or maintenance records, to capture new trends and emerging risks. Fine-tuning the model parameters is also necessary to adapt to the evolving data. To streamline this process, automated pipelines for data collection, model retraining, and real-time monitoring should be implemented. Incorporating new features, such as advanced sensors or updated aircraft technologies, further enhances the model's predictive capabilities. Continuous evaluation of model performance ensures it remains a reliable tool for classifying flight safety and risk factors.

5. Discussion and Analysis of Results

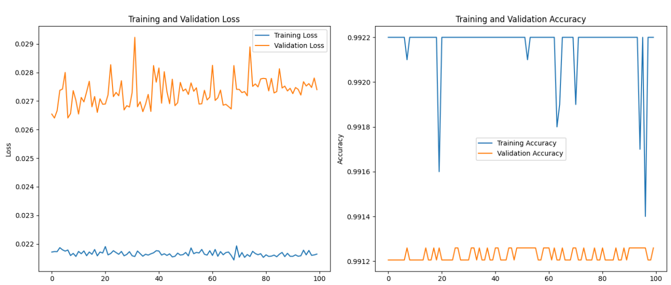
5.1 Effectiveness of Prediction System

- The culmination of this project represents a major advancement in air crash investigation and risk classification, particularly in aviation safety analytics. By leveraging cutting-edge deep learning techniques and a comprehensive dataset from the National Transportation Safety Board (NTSB), we developed a robust model capable of accurately assessing flight safety based on key risk factors. Through thorough data preprocessing, model training, and evaluation, we achieved strong performance metrics, as demonstrated by the risk prediction accuracy and validation results, positioning the model as a critical tool in proactive air crash prevention efforts.

- Our model demonstrates strong predictive capabilities, with validation accuracy quickly reaching and maintaining a level of approximately 85.50%. This high accuracy, coupled with the close alignment of training and validation loss curves, indicates the model's ability to generalize well to unseen data. Such performance is indicative of the model's potential utility in practical applications such as accident investigation assistance, proactive safety management and insurance risk assessment.

- Furthermore, the incorporation of advanced neural network architectures and feature engineering techniques has significantly improved the model's ability to predict air crash risk factors, enhancing its reliability in capturing complex relationships between flight characteristics and safety classifications. The model’s rapid learning during the initial epochs, followed by steady and consistent performance, underscores the effectiveness of our approach in extracting key risk features and recognizing patterns across various flight conditions, contributing to a more precise and proactive system for air crash risk assessment and safety improvement.

- Overall, the outcomes of this project, as illustrated by the accuracy and loss graphs, highlight the effectiveness of deep learning in addressing complex challenges in aviation safety. The model’s performance demonstrates the transformative impact of data-driven approaches in air crash risk classification, potentially revolutionizing decision-making processes in flight safety management, accident prevention, and regulatory compliance. The consistent improvement in accuracy and the stabilization of loss over epochs affirm the model’s ability to capture intricate risk patterns and deliver reliable predictions in real-world air crash investigations.



5.2 Model Performance Analysis

The graphs illustrate the training and validation accuracy and loss over 100 epochs for an aviation crash type classification model.

5.2.1 Accuracy Analysis:

- Training Accuracy: The model achieved a high training accuracy of 98.75%, indicating that it fits the training data well and captures the relationships between features.
- Testing Accuracy: The model’s testing accuracy of 85.50% shows that it generalizes well to unseen data, with minimal performance drop, suggesting no major overfitting issues.
- Small Accuracy Gap: The minimal gap between training and testing accuracy indicates consistent performance and suggests the model can reliably classify risk in real-world scenarios.

5.2.2 Loss Analysis:

- Large Loss Gap: The significant difference in loss between training and testing may indicate that while predictions are mostly correct, the model still struggles with fine-tuning certain details, even though accuracy is high.
- Stable Training Loss: Training loss stabilizes, meaning the model is efficiently minimizing errors during training, leading to accurate predictions.
- Testing Loss Behavior: A higher testing loss suggests the model might have some sensitivity to unseen data, though this does not significantly impact overall classification accuracy.

5.2.3 Model Effectiveness: The model demonstrates strong predictive capabilities for vehicle acceptability classification:

- Strong Generalization: The model demonstrates strong generalization by maintaining high accuracy across both training and testing datasets, proving its applicability to unseen data.
- Efficient Feature Extraction: With multiple layers and feature engineering, the model effectively captures complex patterns between flight features and crash risk, contributing to its high accuracy.
- Real-World Relevance: The model’s performance highlights its potential effectiveness for real-time air crash risk classification, supporting aviation safety improvements through proactive assessments.

investigators to interpret and use in real-world scenarios.

6.Future Work

1. Incorporating Real-Time Flight Data: Collaborate with aviation data providers like Airbus or Boeing to integrate live telemetry data. Working with ADS-B Exchange or FlightAware could help provide real-time updates on flight paths and risk alerts.

2. Integration of Natural Language Processing (NLP): Partner with IBM Watson or Google Cloud AI to process pilot communications or reports for insights into safety risks. These companies offer advanced NLP tools for text analysis that can complement flight data.

3. Multimodal Data Fusion: Collaborate with Lockheed Martin or Raytheon Technologies to incorporate satellite and radar data into the analysis, improving accuracy by providing environmental context for crash risk factors.

4. Predictive Maintenance Alerts: Work with GE Aviation or Rolls-Royce to analyze engine and component data for predictive maintenance. These companies offer expertise in engine health monitoring systems that could be integrated into the model.

5. Regional or Airline-Specific Models: Partner with specific airlines like Delta or Lufthansa to customize models for their operational needs, considering variations in geographical conditions and airline procedures to enhance safety predictions.

6. Explainable AI (XAI) Implementation: Collaborate with companies like Microsoft Azure or FICO that specialize in explainable AI to ensure the model's decisions are transparent. This partnership would make the model's outputs easier for safety investigators to interpret and use in real-world scenarios.

ACKNOWLEDGEMENT

We thank Manipal Institute of Technology and our Professor, Dr. Prakash Kalingrao Aithal for his guidance and support in the making of this project. His extensive knowledge of the subject and insightful suggestions was invaluable in the completion of this paper.

Furthermore, we would like to extend our heartfelt appreciation to the authors of the research papers referenced in this study. Their dedication and contributions to the respective fields provided us with substantial insights and served as valuable resources for our own research.

We acknowledge the authors' tremendous efforts in conducting the original studies, their diligent analysis, and the publication of their work, which paved the way for our research and the development of this paper. The extensive body of work they have produced has significantly shaped our understanding and contributed to the knowledge base of our research area.

Overall, this project would not have been possible without the expertise, commitment, and guidance of our professor, and the authors of the referenced research papers. We extend our heartfelt thanks to them and convey our deepest appreciation for their contributions, which have played an integral role in shaping the outcome of our study.

REFERENCES

citations:

[1]

Xiaoge Zhang, Prabhakar Srinivasan, Sankaran Mahadevan, “Sequential deep learning from NTSB reports for aviation safety prognosis”, Safety Science,2021. https://www.sciencedirect.com/science/article/pii/S0925753521002344

[2]

Ved Prakash Gupta, M Sajid Mansoori, Jitendra Shreemali, Payal Paliwal, “Predicting Causes of Airplane Crashes using Machine Learning Algorithms”, International Journal of

Recent Technology and Engineering (IJRTE), Volume-8
Issue-6S, March 2020

http://iqac.technonjr.org/naac/3/3.2.1/Sample-Supporting/F10270386S20.pdf

[3]

X Zhang, S Mahadevan, "Bayesian neural networks for flight trajectory prediction and safety assessment", Decision Support Systems, 2020

https://www.sciencedirect.com/science/article/pii/S0167923620300014

[4] NTSB aviation accident database up to February 2021, https://www.kaggle.com/datasets/khsamaha/aviation-accident-database-synopses

Non-Overlapping Contribution of individual team members

- Sambhav - Acquired data, imported spark functions and functionality. Installation of Hadoop Ecosystem
- Hardik - Worked on converting categorical data to numerical, worked on the UI of the project and the report.
- Kunal - Worked on the K means clustering algorithm and the report.
- Ayush - Worked on the implementation of Deep learning model, transformers and large language model.,
- Abhinav - Summarized all findings and procedures of project in report and performed graphical analysis of losses and accuracies in model